Akhil Reddy Edla

Evansville, IN | +1(913) 327 - 0543 | akhilreddyedla99@gmail.com | LinkedIn

PROFESSIONAL SUMMARY

Data Engineer with 5+ years of experience designing, building, and optimizing end-to-end **data pipelines**, **lakehouse** architectures, and cloud analytics platforms on **AWS** and **GCP**. Skilled in real-time and batch data processing using **Spark**, **PySpark**, **Kafka**, **Pub/Sub**, **SQL**, and **Python**, handling millions to billions of records efficiently. Experienced in ETL orchestration with **Airflow**, **CI/CD deployment**, and **DevOps** practices, ensuring scalable, reliable, and maintainable workflows. Adept at collaborating with business and analytics teams to deliver analytics-ready datasets, **BI dashboards**, and actionable insights, driving data-driven decision-making and operational efficiency.

TECHNICAL SKILLS

Data Engineering & ETL: Airflow, dbt, Apache NiFi, Talend, AWS Glue, Cloud Data Pipelines, Data Lake/Lakehouse design, ETL orchestration, Event-driven pipelines.
Big Data & Streaming: Spark (PySpark, Spark SQL, Spark Streaming), Hadoop, Hive, Kafka, Delta Lake, Flink, Snowflake.
Cloud Platforms: AWS (S3, Glue, Redshift, Lambda, EMR, EC2), GCP (BigQuery, Cloud Functions, Cloud Run, DataStream, Pub/Sub, Composer), Azure (Data Factory, Synapse, Blob Storage).
Programming & Scripting: Python (Pandas, NumPy, PySpark), SQL, Shell Scripting, Java.
Databases & Warehouses: PostgreSQL, SQL Server, Oracle, BigQuery, Snowflake, Redshift, MySQL.
DevOps & Infrastructure: Docker, Kubernetes, Terraform, Jenkins, Git/GitHub, CI/CD pipelines, Cloud Monitoring and Logging.
DataOps & Observability: Data Quality frameworks, Logging & Monitoring pipelines, Metadata management, Data Governance.

PROFESSIONAL EXPERIENCE

Data Engineer **Oct 2024 - Present**
CenterPoint Energy | Evansville, IN

- Built and optimized ETL pipelines using **Python**, **PySpark**, and **SQL** to process structured and unstructured datasets, enhancing analytics query speed by 40% for business teams.
- Leveraged GCP **Pub/Sub**, **Cloud Functions**, **Cloud Run**, and **DataStream** to stream millions of events daily, enabling near real-time insights for enterprise dashboards.
- Automated replication of 150 Coupa tables from **Snowflake** to **BigQuery**, ensuring reliable daily updates while cutting manual refresh work by 70%.
- Integrated SAP **SLT** workflows with GCP infrastructure to maintain synchronized enterprise data, improving consistency across downstream reporting systems.
- Developed Adobe Analytics ingestion pipelines into **GCS** and **BigQuery**, accelerating marketing and product reporting with real-time visibility.
- Created **Power BI** dashboards to **monitor pipeline** performance and data quality, strengthening transparency and trust in key datasets.
- Partnered with business teams to optimize workflows and data models, delivering actionable datasets that reduced reporting delays by 30%.
- Implemented CI/CD pipelines using GitHub, Cloud Build, and Terraform, streamlining deployments and reducing production errors by 60%.
- Utilized **Atlan** to maintain and enhance **dataset documentation** by adding detailed column descriptions, business context, lineage notes, and tags, improving data clarity for analysts and downstream teams.
- Created and updated **Atlan metadata** for **BigQuery** tables across multiple pipelines, ensuring consistent naming standards, ownership assignments, and improved data discoverability.
- Leveraged **Monte Carlo** to configure and refine data quality monitors for freshness, volume, and schema changes, strengthening reliability and early detection of anomalies across key datasets.
- Collaborated with data analysts to validate dataset definitions, maintain data dictionaries, and ensure **governance standards** were followed across new and existing pipelines.
- Developed data ingestion and preprocessing workflows for an **AI pipeline** designed to **detect potential pipe corrosion** around gas meters using prompt-based LLM techniques and historical dataset patterns.
- **Maintained comprehensive documentation** and **knowledge transfer**, reducing onboarding time for new engineers by 50%.

Data Engineer **Dec 2022 - Oct 2024**
Fifth Third Bank | Evansville, IN

- Engineered batch and streaming ETL pipelines on AWS using Python, PySpark, and **Airflow**, improving data availability for analytics by 45%.
- Developed Spark Structured Streaming pipelines with **Kafka**, enabling real-time **fraud detection** and cutting incident response times in half.
- Automated **data validation** processes with Python and AI-assisted CoPilot, increasing pipeline reliability and throughput by 70%.
- Designed and executed POCs for fraud and opioid analytics, demonstrating measurable operational improvements through real-time PySpark workflows.
- Migrated on-premises SQL Server datasets to **S3** and **Redshift**, reducing query runtimes by 50% and supporting scalable analytics pipelines.

- Exposed curated datasets via APIs, standardizing data delivery and enabling faster cross-team analytics consumption.
- Orchestrated **Airflow DAGs** for scheduled and event-driven workflows, reducing pipeline failures by 40% while improving operational efficiency.
- Implemented proactive **monitoring** and **alerting** frameworks, achieving over 99% pipeline uptime and ensuring reliable reporting.

Data Engineer **Jul 2020 - Dec 2022**

Infogain | Bangalore, India

- Designed high-frequency ETL pipelines using Python multiprocessing, **Kafka**, **Spark**, and **Airflow**, processing millions of API calls daily and lowering ingestion latency by 60%.
- Built **event-driven** frameworks with Spark Structured Streaming to make real-time data immediately available for business dashboards.
- Applied Pytest-based automation for data validation, raising reporting accuracy to 99% and reducing manual data checks.
- Streamlined **BigQuery** and **Snowflake** data models, accelerating report generation by 40% and improving analytics delivery.
- Containerized ETL pipelines with Docker and deployed on GCP, increasing workflow scalability and maintainability.
- Developed batch and streaming pipelines integrating **GCS**, **BigQuery**, and **Snowflake**, enabling rapid access to high-quality datasets.
- Optimized Spark jobs and SQL queries to improve pipeline efficiency by 40%, lowering overall cloud resource costs.

PROJECTS

Snowflake to BigQuery Migration

- Executed migration of enterprise datasets from Snowflake to BigQuery using Cloud Run batch jobs, optimizing query performance and reducing operational costs.
- Implemented schema mapping, incremental data loads, and validation checks to ensure data accuracy and consistency across platforms.
- Leveraged SQL, Python, and Google Cloud monitoring tools to automate workflows and track data migration progress efficiently.

SAP Data Ingestion Pipeline

- Built a scalable ingestion pipeline to transfer SAP HANA data into BigQuery using SQL queries provided by the reporting team.
- Designed ETL transformations with Python and Cloud Composer to harmonize data for reporting and analytics needs.
- Scheduled automated data loads and implemented error-handling mechanisms to maintain reliability and timeliness.

SQL Server to BigQuery Replication

- Developed a data replication pipeline from SQL Server to BigQuery using DataStream for low-latency analytics.
- Monitored streaming pipelines with Cloud Monitoring and implemented alerting for anomalies to ensure continuous data availability.

Adobe Analytics Data Automation

- Automated ingestion of Adobe Analytics data feeds into BigQuery using Python and GCS integration to enable advanced reporting.
- Developed data cleaning, transformation, and aggregation workflows for digital marketing insights.
- Scheduled pipelines using Cloud Scheduler and monitored ETL processes to ensure timely delivery of analytics-ready datasets.

Event-Driven GCS to BigQuery Ingestion

- Created a Cloud Function triggered by Eventarc to load files from Google Cloud Storage into BigQuery in real time.
- Implemented dynamic table creation, data validation, and logging to handle diverse file formats and ensure accuracy.
- Leveraged Python, BigQuery APIs, and cloud-native event triggers to streamline ingestion and reduce latency for analytics teams.

EDUCATION

Master of Science in Computer Science

University of Central Missouri | Warrensburg , MO

Bachelor of Technology in Computer Science

Jawaharlal Nehru Technological University | Hyderabad , India

---

CERTIFICATIONS

- GCP Data Engineer Professional Certification
- Cloud Practitioner - **AWS**
- Scrum/Agile Methodology Certification
- Data Engineering with Python **- Coursera**
- Big Data Analytics Specialization **- Coursera**
- Power BI for Data Analysts **- LinkedIn Learning**